ERGA Assembly Report

v24.10.15

Tags: ATLASea[INVALID TAG]

TxID	1529433
ToLID	**ecDavRubi1**
Species	Davidaster rubiginosus
Class	Crinoidea
Order	Comatulida

Genome Traits	Expected	Observed
Haploid size (bp)	636,929,079	689,625,113
Haploid Number	4 (source: ancestor)	14
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.7.Q51

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

. Observed Haploid Number is different from Expected

. Kmer completeness value is less than 90 for collapsed
. Not 90% of assembly in chromosomes for collapsed

Curator notes

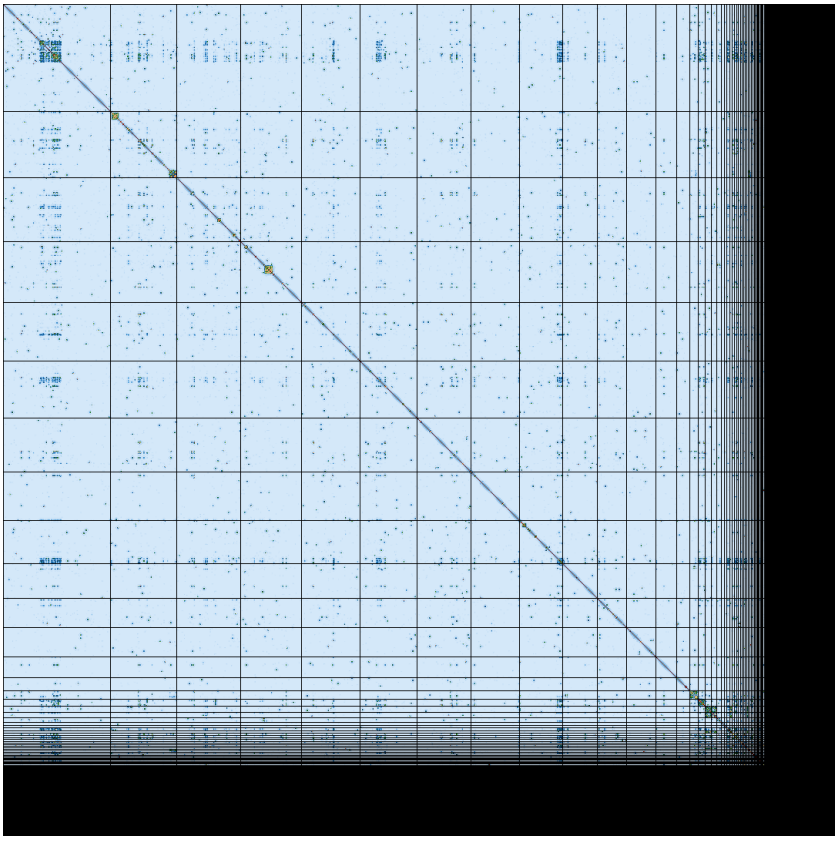
. Interventions/Gb: 9
. Contamination notes: ""
. Other observations: "The assembly of Davidaster rubiginosus (ecDavRubi1) is based on 30X PacBio data and 130X Arima Hi-C data generated as part of the ATLASea programme (https://www.atlasea.fr). The assembly process included the following steps: initial PacBio assembly generation with Hifiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 24 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 0.878 Mb (with the largest being 0.126 Mb). Additionally, 218 regions totaling 51.141 Mb (with the largest being 1.463 Mb) were identified as haplotypic duplications and removed. The mitochondrial genome was assembled using Oatk. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 2 haplotypic regions were removed, totaling 1.166Mb (with the largest being 0.624Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

Metrics	Pre-curation collapsed	Curated collapsed
Total bp	690,802,779	689,625,113
GC %	33.37	33.37
Gaps/Gbp	247.54	246.51
Total gap bp	17,100	17,200
Scaffolds	838	837
Scaffold N50	44,496,015	44,496,015
Scaffold L50	7	7
Scaffold L90	35	35
Contigs	1,009	1,007
Contig N50	6,486,000	6,486,000
Contig L50	30	30
Contig L90	157	156
QV	51.4326	51.4286
Kmer compl.	81.7269	81.6519
BUSCO sing.	93.5%	92.6%
BUSCO dupl.	4.3%	3.7%
BUSCO frag.	1.2%	2.5%
BUSCO miss.	1.0%	1.2%

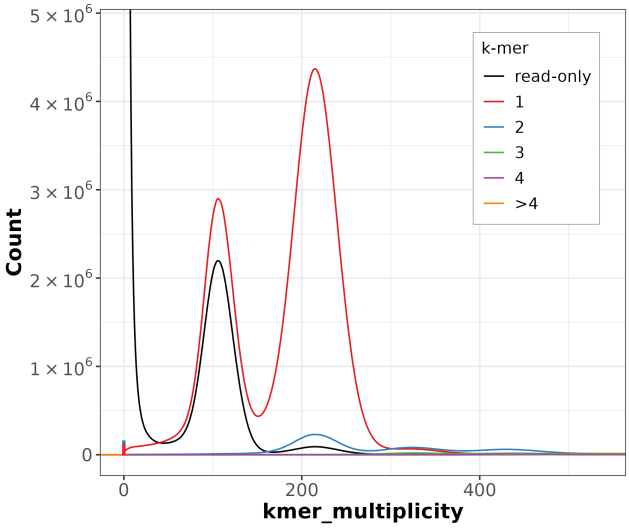
Warning! BUSCO versions or lineage datasets are not the same across results:
BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: metazoa_odb12 (genomes:206, BUSCOs:672)

HiC contact map of curated assembly

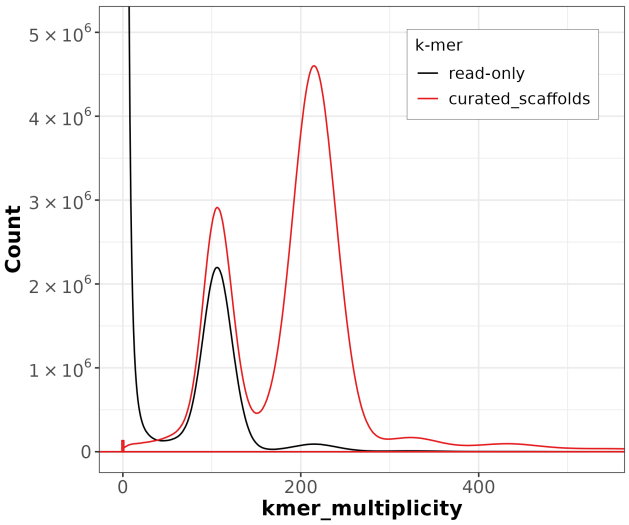


collapsed [LINK]

K-mer spectra of curated assembly

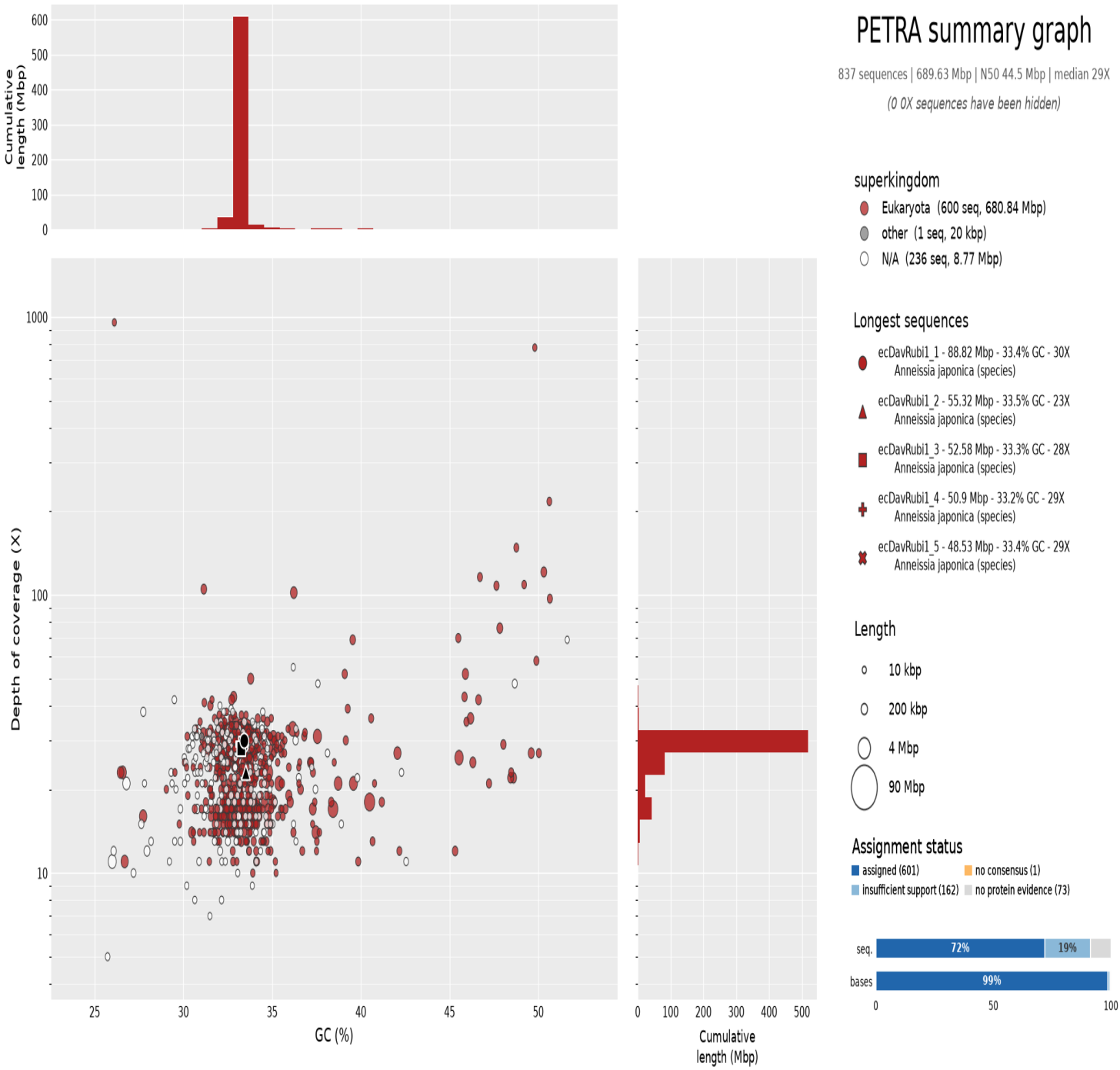


Distribution of k-mer counts per copy numbers found in asm



Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	Long reads	Arima
Coverage	30	130

Assembly pipeline

- **Hifiasm**
  |_ *ver:* 0.19.5-r593
  |_ *key param:* NA
- **purge_dups**
  |_ *ver:* 1.2.5
  |_ *key param:* NA
- **YaHS**
  |_ *ver:* 1.2
  |_ *key param:* NA

Curation pipeline

- **PretextMap**
  |_ *ver:* 0.2.4
  |_ *key param:* NA
- **PretextView**
  |_ *ver:* 1.0.7
  |_ *key param:* NA

Submitter: Arnaud Couloux
Affiliation: Genoscope

Date and time: 2026-10-05 19:14:10 CEST